基于对偶图正则化的多层概念分解算法 *

张显 叶军

(南京邮电大学 理学院，南京 210023)

摘　要：摘　要：为了进一步挖掘数据间的隐藏信息，在多层概念分解(MCF)算法的框架下，考虑每一层分解下的数据流形和特征流形，提出了一种基于对偶图正则化的多层概念分解(DGMCF)算法。该算法通过对数据的逐层分解，以分层的方式学习，并在每一层分解数据中构建数据空间和特征属性空间的拉普拉斯图，用于反映数据流形和特征流形的多元几何结构信息，从而能够更好地从复杂数据中提取出更有效的特征。采用交替迭代的方法求解算法的目标函数并证明了算法的收敛性。通过在三个真实数据库(TDT2、PIE、COIL20) 上的实验表明，该方法在数据的聚类表示效果方面优于其他方法。

关键词：概念分解；多层分解；对偶回归；流形学习；聚类

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.10.0935

Dual-graph regularized multilayer concept factorization

Zhang Xian, Ye Jun

(*School of Natural Sciences*, *Nanjing University of Posts & Telecommunications*, *Nanjing* 210023, *China*)

Abstract: In order to further excavate the hidden information between data, under the framework of multilayer concept factorization (MCF) algorithm, this paper proposed a novel algorithm called dual-graph regularized multilayer concept factorization (DGMCF) algorithm, which encoded the geometric structure information of data and feature spaces by constructing two Laplacian regularize term in each layer factorization, respectively. By this way, the proposed method could learn features in a hierarchical manner, and thus provided a better chance for learning meaningful features from the complex data. Moreover, it developed the iterative updating optimization scheme for DHCF, and also provided the convergence proof of the optimization scheme. Experimental results on TDT2 document datasets, PIE and COIL20 image datasets demonstrate the effectiveness of our proposed method.

Key Words: CF; multilayer factorization; dual regularized; manifold learning; clustering

0　引言

如何从海量的高维数据中挖掘出隐藏信息和有效数据，已成为现今机器学习、数据挖掘、社会网络分析等领域的研究热点[1~3]。合适的数据表示方式能够挖掘出数据中的潜在结构，有利于数据进一步的处理。目前，矩阵分解方法作为一种有效的数据处理方式引起了许多研究者的关注。常用的矩阵分解算法包括奇异值分解(singular value decomposition, SVD)、非负矩阵分解(nonnegative matrix factorization, NMF)[4]和概念分解(concept factorization, CF)[5]等。

在 NMF 和 CF 的框架下，已有许多文献对其进行了拓展。结合标签信息，Liu 等人[6]提出了半监督的矩阵分解方法，一种是约束非负矩阵分解(constrained NMF, CNMF)，另一种是约束概念分解(constrained CF, CCF)[7]。这两种方法都是将标签信息作为硬约束分别嵌入在 NMF 和 CF 框架中的。将流形学习的思想[8~11]融入至 NMF 和 CF 方法中，文献[12]提出了图正则的非负矩阵分解（graph regularized NMF, GNMF）算法，文献[13]提出了局部连续概念分解算法(locally consistent CF, LCCF)，这两种方法在文本聚类及人脸识别等应用中均取得了不错的效果。

然而以上方法均是对数据进行单层分解。已有文献表明[14,15]，对数据进行单层分解很难获取原始数据中隐藏的层次信息，从而不能更好地对数据进行表示。因此，X.Li et al.[16]提出了多层非负矩阵分解(Multilayer NMF, MNMF)算法，该方法能更好的从混合信号中对原信号进行分离。 在 CF 的框架下，Li 等人[16]提出了多层概念分解(multilayer CF, MCF)，通过利用多层模型，MCF 方法能够从高维数据中获取数据隐藏的层次信息。考虑到数据的几何结构信息，最近 Li 等人[17]在 MCF 方法的基础上，在每一层概念分解中结合流形学习的思想，提出了

基金项目：江苏省自然科学基金资助项目（BK20150867)；南京邮电大学国家自然科学基金孵化资助项目（NY215125）
作者简介：张显（1994-），男，硕士研究生，主要研究方向为模式识别、人脸识别、机器学习、计算机视觉；叶军（1981-），男，副教授，博士，主要研究方向为模式识别、机器学习、图像处理（yj8422092@163.com）.

图回归的多层概念分解(graph regularized multilayer CF, GMCF)。GMCF 方法不仅使用了多层结构，而且还考虑了每个数据层的流形几何结构，的确有助于挖掘数据最本质的信息。

GNMF、LCCF 和 GMCF 方法都利用了数据的空间结构信息，它们可以有效地提高学习的质量，然而这些工作只考虑了数据空间的分布结构，没有利用特征属性空间的结构信息。最近， Shang 和 Ye 等人[18,19]分别在非负矩阵分解和概念分解框架下提出了同时考虑了数据流形和特征流形的几何结构的双图正则化非负矩阵分解(Graph Dual regularization Nonnegative Matrix Factorization, DNMF)算法和双图正则化概念分解(graph dual regularization concept factorization, GCF)算法，均取得了不错的效果。这也进一步说明了不但数据空间的几何结构信息可以有效地提高学习的质量，同时特征属性空间的几何结构信息也能对学习的质量起到辅助作用。

因此，为了进一步挖掘数据间的隐藏信息，同时利用好数据的几何结构信息，在多层概念分解(MCF)算法的框架下，考虑每一层分解下的数据流形和特征流形，提出了一种基于对偶图正则化的多层概念分解算法(dual graph regularized multilayer concept factorization，DGMCF)。 该算法在每一层分别在数据空间和特征空间构建图来反映它们各自的分布流形几何结构信息，以此来获得数据的几何结构信息。建立基于对偶图正则化的多层概念分解模型，推导出该算法的交替迭代更新规则，给出该算法的收敛性证明，实验结果表明了算法的有效性和准确性。

1 相关工作

1.1 概念分解(CF)算法

给定非负矩阵 $X=[x_1,\cdots,x_N]\in R^{M\times N}$， X 的每一列代表一个样本。CF 算法的目标是寻求两个非负矩阵 $W=[w_{jk}]\in R^{N\times K}$，$V=[v_{jk}]\in R^{N\times K}$，其中 $K\ll\min\{M,N\}$，使其满足 $X\approx XWV^T$。CF 的目标函数可表示为

$$\min_{W,V}:J_{CF}=\|X-XWV^T\|_F^2\quad s.t. W,V\geq 0 \tag{1}$$

相应的乘积更新迭代规则在文献[5]中已详细地给出。

1.2 多层概念分解(MCF)算法

面对一些复杂数据，特别是病态数据， Li 等人[16]在概念分解的基础上采用分层的方法，逐层地对数据进行概念分解，提出了多层概念分解(MCF)算法。MCF 的思想是：记 $X=XW_1V_1$ 为第 1 层分解，将 V_1 作为初始数据进行第二层分解得 $V_1=V_1W_2V_2$，依此类推，得到第 L 层的概念分解 $V_{L-1}=V_{L-1}W_LV_L$，则最终 $X=XW_1V_1W_2V_2\cdots W_LV_L$，且 $W=W_1V_1W_2V_2\cdots W_L$，$V=V_L$。多层概念分解的目标函数可表示为

$$\min_{W_l,V_l}:J_{MCF}=\|X_l-X_lW_lV_l^T\|_F^2\quad s.t. W_l,V_l\geq 0 \tag{2}$$

相应的乘积更新迭代规则在文献[5]中已详细地给出。

1.3 双图正则化概念分解(GCF)算法

Ye 等人[19]在概念分解算法的基础上同时考虑了数据和特征属性的几何结构信息， 提出了双图正则化的概念分解算法。其目标函数可表示为

$$\begin{aligned}\min_{W,V}:J_{GCF}&=\|X-XWV^T\|^2\\&+\alpha\mathrm{Tr}(V^TL_VV)+\beta\mathrm{Tr}(W^TL_WW)\\ s.t.\ &W,V\geq 0\end{aligned} \tag{3}$$

其中：

$$L_W=X^TL_UX=X^T(D^U-S^U)X=D^W-S^W$$

针对式(3)的乘积更新迭代规则在文献[16]中已详细地给出。

2 基于对偶图正则化的多层概念分解(DGMCF)

2.1 构建对偶图正则项

最近的研究表明：不仅观测到的数据分布在一个低维子流形上，称之为数据流形；而且数据的特征也分布在一个低维子流形上，称之为特征流形[18]。在对数据进行概念分解的过程中，由于每一层数据分解后均涉及数据流形及特征流形的几何结构信息获取，所以在每一层分别用两个图来刻画数据流形和特征流形的几何结构，即数据图和特征图。

第 l 层分解下数据图构造

设第 l 层分解下图的顶点集合为数据集 $\{(\mathbf{x}_1)_l,\cdots,(\mathbf{x}_N)_l\}$，若第 j 个样本与第 s 个样本互为近邻点，则第 j 个样本与第 s 个样本之间存在一条边，其权值为 $(s_{js}^V)_l$， 对应的邻域矩阵：

$$(\mathbf{S}_{js}^V)_l=\begin{cases}1, & if\ (\mathbf{x}_s)_l\in N_p((\mathbf{x}_j)_l);\\ 0, & otherwise.\end{cases}\quad j,s=1,\cdots,N \tag{4}$$

其中： $N_p((\mathbf{x}_j)_l)$ 表示第 l 层分解下 $\mathbf{x}_j$ 的 p 最近邻数据样本集。

此时在第 l 层分解下数据图的拉普拉斯矩阵可表示为 $(L_V)_l=D_l^V-S_l^V$。令 $V_l=[(v_1^T)_l,\cdots,(v_N^T)_l]^T$ 为待求的第 l 层分解下的低维数据表示，则其表示的平滑度为

$$\begin{aligned}(\mathcal{O})_l&=\frac{1}{2}\sum_{j,s=1}^{N}(\mathbf{S}_{js}^V)_l\|(v_j)_l-(v_s)_l\|^2\\&=\mathrm{Tr}(V_l^TD_l^VV_l)-\mathrm{Tr}(V_l^TS_l^VV_l)\\&=\mathrm{Tr}(V_l^T(L_V)_lV_l)\end{aligned} \tag{5}$$

第 l 层分解下特征图构造

类似地，在第 l 层分解下将图的顶点集合定义为特征集 $\{(\mathbf{x}_1^T)_l,\cdots,(\mathbf{x}_M^T)_l\}$， 对应的领域矩阵定义如下：

$$(\mathbf{S}_{js}^W)_l=\begin{cases}1, & if\ (\mathbf{x}_s^T)_l\in N_p((\mathbf{x}_j^T)_l);\\ 0, & otherwise.\end{cases}\quad j,s=1,\cdots,M \tag{6}$$

可得到在第 l 层分解下特征图的拉普拉斯矩阵表示为 $(L_W)_l=X_l^T(L_U)_lX_l=X_l^T(D_l^U-S_l^U)X_l=D_l^V-S_l^V$ ， 令

$W_l = \left[(w_1^T)_l, \cdots, (w_M^T)_l\right]^T$ 为待求的低维数据表示，则相应的平滑度为

$$\begin{aligned}(\mathcal{O}_2)_l &= \frac{1}{2}\sum_{j,s=1}^{M}(S_{js}^W)_l \left\|(w_j^T)_l - (w_s^T)_l\right\|^2 \\ &= \mathrm{Tr}(W_l^T D_l^W W_l) - \mathrm{Tr}(W_l^T S_l^W W_l) \\ &= \mathrm{Tr}(W_l^T (L_W)_l W_l)\end{aligned} \tag{7}$$

2.2 构建 DGMCF 算法的目标函数

为了同时考虑样本的数据流形和特征流形的几何结构信息，在 MCF 算法的目标函数中添加基于数据图和特征图的正则项，得到 DGMCF 算法的目标函数为

$$\begin{aligned}&\min_{W_l, V_l} : J_{DGMCF} = \left\|X_l - X_l W_l V_l^T\right\|^2 + \alpha(\mathcal{O}_1)_l + \beta(\mathcal{O}_2)_l \\ &\text{s.t.} \quad W_l, V_l \ge 0\end{aligned} \tag{8}$$

其中：$\alpha \ge 0, \beta \ge 0$ 为正则化参数。

2.3 DGMCF 目标函数的求解

DGMCF 算法中的目标函数 J_{DGMCF} 是关于两个变量 W_l 和 V_l 的非凸函数，因此求其全局最优解是不现实的。利用交替迭代法可得到问题的局部最优解。记 $K_l = X_l^T X_l$，则目标函数 J_{DGMCF} 可重写为

$$\begin{aligned}J_{DGMCF} &= \mathrm{Tr}[(X_l - X_l W_l V_l^T)^{\mathrm{T}}(X_l - X_l W_l V_l^T)] \\ &+\alpha \mathrm{Tr}\left(V_l^T (L_V)_l V_l\right) + \beta \mathrm{Tr}\left(W_l^T (L_W)_l W_l\right) \\ &= \mathrm{Tr}\left(K_l - 2V_l W_l^T K_l + V_l W_l^T K_l W_l V_l^T\right) \\ &+\alpha \mathrm{Tr}\left(V_l^T (L_V)_l V_l\right) + \beta \mathrm{Tr}\left(W_l^T (L_W)_l W_l\right)\end{aligned} \tag{9}$$

令 $\boldsymbol{\Psi}_l = [\psi_{jk}^l]$，$\boldsymbol{\Phi}_l = [\varphi_{jk}^l]$ 为约束 $W_l \ge 0$ 和 $V_l \ge 0$ 对应的拉格朗日乘子，则式(9)的拉格朗日函数 L 为

$$\begin{aligned}L &= \mathrm{Tr}\left(K_l - 2V_l W_l^T K_l + V_l W_l^T K_l W_l V_l^T\right) \\ &+ \alpha \mathrm{Tr}\left(V_l^T (L_V)_l V_l\right) + \beta \mathrm{Tr}\left(W_l^T (L_W)_l W_l\right) \\ &+ \mathrm{Tr}(\boldsymbol{\Psi}_l W_l^T) + \mathrm{Tr}(\boldsymbol{\Phi}_l V_l)\end{aligned} \tag{10}$$

对函数 L 分别关于 W_l 和 V_l 求偏导，由 KKT 最优性条件可以得到 DGMCF 算法的更新迭代公式为

$$(w_l)_{jk}^{(t+1)} \leftarrow (w_l)_{jk}^{(t)} \frac{(K_l V_l + \beta(S^W)_l W_l)_{jk}}{(K_l W_l V_l^T V_l + \beta(D^W)_l W_l)_{jk}} \tag{11}$$

$$(v_l)_{jk}^{(t+1)} \leftarrow (v_l)_{jk}^{(t)} \frac{(\mathbf{K}_l \mathbf{W}_l + \alpha(\mathbf{S}^V)_l \mathbf{V}_l)_{jk}}{(\mathbf{V}_l \mathbf{W}_l^T \mathbf{K}_l \mathbf{W}_l + \alpha(\mathbf{D}^V)_l \mathbf{V}_l)_{jk}} \tag{12}$$

DGMCF 方法将在迭代次数达到最大时停止或在停止准则 $\left\|J_{DGMCF}^t - J_{DGMCF}^{t-1}\right\| < \varepsilon$ 满足时停止迭代。

2.4 DGMCF 算法的收敛性证明

定义 1 当满足条件：$G(w,w') \ge F(w)$ 和 $G(w,w) = F(w)$ 时，$G(w,w')$ 为 $F(w)$ 的一个辅助函数。

引理 1 若 G 为 F 的辅助函数，则函数 F 在如下的更新公式下为单调下降的。

$$w^{(K+1)} = \arg\min_w G(w, w^{(K)}) \tag{13}$$

证明

$$F(w^{(K+1)}) \le G(w^{(K+1)}, w^{(K)}) \le G(w^{(K)}, w^{(K)}) = F(w^{(K)}) \cdot$$

令 $(w_l)_{ab}$ 为矩阵 W_l 的元素，$F_{(w_l)_{ab}}$ 为目标函数 J_{DGMCF} 中仅与 $(w_l)_{ab}$ 元素有关的项，即

$$F_{(w_l)_{ab}} = \left\|X_l - X_l W_l V_l^T\right\|^2 + \beta \mathrm{Tr}\left(W_l^T (L_W)_l W_l\right) \tag{14}$$

考虑到算法是基于元素运算的，故首先证明 $F_{(w_l)_{ab}}$ 在式(11)下为单调下降的。事实上由于：

$$F'_{(w_l)_{ab}} = \left(-2K_l V_l + 2K_l W_l V_l^T V_l + 2\beta (L_W)_l W_l\right)_{ab}$$

$$F''_{(w_l)_{ab}} = 2(K_l)_{aa}\left(V_l^T V_l\right)_{bb} + 2\beta\left((L_W)_l\right)_{aa}$$

引理 2 函数 $G(w_l,(w_l)_{ab}^{(t)})$ 为 $F_{(w_l)_{ab}}$ 的辅助函数。

$$\begin{aligned}&G(w_l,(w_l)_{ab}^{(t)}) = F_{(w_l)_{ab}}((w_l)_{ab}^{(t)}) \\ &+F'_{(w_l)_{ab}}((w_l)_{ab}^{(t)})(w_l - (w_l)_{ab}^{(t)}) \\ &+\frac{(K_l W_l V_l^T V_l)_{ab} + \beta((D^W)_l W_l)_{ab}}{(w_l)_{ab}^{(t)}}(w_l - (w_l)_{ab}^{(t)})^2\end{aligned} \tag{15}$$

证明 由定义 1，显然 $G(w_l, w_l) = F_{(w_l)_{ab}}(w_l)$。令 $F_{(w_l)_{ab}}(w_l)$ 的 Taylor 展开序列为

$$\begin{aligned}F_{(w_l)_{ab}}(w_l) &= F_{(w_l)_{ab}}((w_l)_{ab}^{(t)}) + F'_{(w_l)_{ab}}((w_l)_{ab}^{(t)})(w_l - (w_l)_{ab}^{(t)}) \\ &+[(K_l)_{aa}\left(V_l^T V_l\right)_{bb} + \beta\left((L_W)_l\right)_{ab}](w_l - (w_l)_{ab}^{(t)})^2\end{aligned}$$

由式(15)知，证明 $G(w_l,(w_l)_{ab}^{(t)}) \ge F_{(w_l)_{ab}}(w_l)$ 等价于证明：

$$\begin{aligned}&\frac{\left(K_l W_l V_l^T V_l\right)_{ab} + \beta((D^W)_l W_l)_{ab}}{(w_l)_{ab}^{(t)}} \\ &\ge (K_l)_{aa}(V_l^T V_l)_{bb} + \beta\left((L_W)_l\right)_{aa}\end{aligned} \tag{16}$$

事实上：

$$(K_l W_l V_l^T V_l)_{ab} = \sum_{i=1}^{k}(K_l W_l)_{ai}\left(V_l^T V_l\right)_{ib} \ge (K_l W_l)_{ab}\left(V_l^T V_l\right)_{bb}$$

$$\ge \sum_{i=1}^{k}(K_l)_{ai}(w_l)_{ib}^{(t)}\left(V_l^T V_l\right)_{bb} \ge (w_l)_{ab}^{(t)}(K_l)_{aa}\left(V_l^T V_l\right)_{bb}$$

且

$$\beta((D^W)_l W_l)_{ab} = \beta\sum_{j=1}^{M}((D^W)_l)_{aj}(w_l)_{jb}^{(t)} \ge \beta((D^W)_l)_{aa}(w_l)_{ab}^{(t)}$$

$$\ge \beta((D^W)_l - (S^W)_l)_{aa}(w_l)_{ab}^{(t)} = \beta((L_W)_l)_{aa}(w_l)_{ab}^{(t)}$$

因此，可知式(15)成立，即 $G(w_l,(w_l)_{ab}^{(t)}) \ge F_{(w_l)_{ab}}(w_l)$。

引理 3 函数 $G(v_l,(v_l)_{ab}^{(t)})$ 为 $F_{(v_l)_{ab}}$ 的辅助函数。

$$G(v_l,(v_l)_{ab}^{(t)})=F_{(v_l)_{ab}}((v_l)_{ab}^{(t)})+F'_{v_{ab}}((v_l)_{ab}^{(t)})(v_l-(v_l)_{ab}^{(t)}) \\ +\frac{(\boldsymbol{V}_l\boldsymbol{W}_l^T\boldsymbol{K}_l\boldsymbol{W}_l)_{ab}+\alpha((\boldsymbol{D}^V)_l\boldsymbol{V}_l)_{ab}}{(v_l)_{ab}^{(t)}}(v_l-(v_l)_{ab}^{(t)})^2 \quad (17)$$

其中：$F_{(v_l)_{ab}}=\left\|\boldsymbol{X}_l-\boldsymbol{X}_l\boldsymbol{W}_l\boldsymbol{V}_l^T\right\|^2+\alpha\operatorname{Tr}(\boldsymbol{V}_l^T(\boldsymbol{L}_V)_l\boldsymbol{V}_l)$。

引理 3 的证明过程同引理 2 的证明，限于篇幅，此处具体证明参见引理 2。

定理 1 对于给定的数据 $\boldsymbol{X}$ 及任意的 $\boldsymbol{W}_l\geq 0,\ \boldsymbol{V}_l\geq 0$，提出的交替迭代更新规则式 (11) 和 (12) 可使得目标函数 $\boldsymbol{J}_{DGMCF}(\boldsymbol{W}_l,\boldsymbol{V}_l)$ 单调下降。

证明 由引理 2,3，将式(15) (17)分别代入式(13)得

$$(w_l)_{ab}^{(t+1)}=(w_l)_{ab}^{(t)}-(w_l)_{ab}^{(t)}\frac{F'_{(w_l)_{ab}}((w_l)_{ab}^{(t)})}{2(\boldsymbol{K}_l\boldsymbol{W}_l\boldsymbol{V}_l^T\boldsymbol{V}_l)_{ab}+2\beta((\boldsymbol{D}^W)_l\boldsymbol{W}_l)_{ab}} \\ =(w_l)_{ab}^{(t)}\frac{(\boldsymbol{K}_l\boldsymbol{V}_l+\beta(\boldsymbol{S}^W)_l\boldsymbol{W}_l)_{ab}}{(\boldsymbol{K}_l\boldsymbol{W}_l\boldsymbol{V}_l^T\boldsymbol{V}_l+\beta(\boldsymbol{D}^W)_l\boldsymbol{W}_l)_{ab}} \quad (18)$$

$$(v_l)_{ab}^{(t+1)}=(v_l)_{ab}^{(t)}-(v_l)_{ab}^{(t)}\frac{F'_{(v_l)_{ab}}((v_l)_{ab}^{(t)})}{2(\boldsymbol{V}_l\boldsymbol{W}_l^T\boldsymbol{K}_l\boldsymbol{W}_l)_{ab}+2\alpha((\boldsymbol{D}^V)_l\boldsymbol{V}_l)_{ab}} \\ =(v_l)_{ab}^{(t)}\frac{(\boldsymbol{K}_l\boldsymbol{W}_l+\alpha(\boldsymbol{S}^V)_l\boldsymbol{V}_l)_{ab}}{(\boldsymbol{V}_l\boldsymbol{W}_l^T\boldsymbol{K}_l\boldsymbol{W}_l+\alpha(\boldsymbol{D}^V)_l\boldsymbol{V}_l)_{ab}} \quad (19)$$

由于式(15) (17)分别为 $F_{(w_l)_{ab}}$ 及 $F_{(v_l)_{ab}}$ 的辅助函数，所以 $F_{(w_l)_{ab}}$ 及 $F_{(v_l)_{ab}}$ 分别在迭代更新式(11)(12)下为单调下降。

2.5 DGMCF 算法的具体步骤

输入：数据集矩阵 $\boldsymbol{X}$。

1：初始化参数。设定最大迭代次数(IterMax)及误差限 ε，设定近邻点参数 p、分解维度 K、正则化参数 α 和 β。

2：记 $\boldsymbol{X}_1=\boldsymbol{X}$，并随机生成非负矩阵 $\boldsymbol{W}$ 和 $\boldsymbol{V}$。

for l =1: L

for t =1: IterMax (t 为迭代次数)

3：计算数据图和特征图的邻接矩阵 $(S^V)_l$ 和 $(S^W)_l$。

4: 计算对角矩阵 $(D^V)_l$ 和 $(D^W)_l$。

5：固定 $\boldsymbol{V}_l^{(t)}$，根据式(10)更新 $\boldsymbol{W}_l$ 得 $\boldsymbol{W}_l^{(t+1)}$。

6：固定 $\boldsymbol{W}_l^{(t+1)}$，根据式(11)更新 $\boldsymbol{V}_l$ 得 $\boldsymbol{V}_l^{(t+1)}$。

7:重复步骤 2 和 3,直至终止条件 $\left\|\boldsymbol{J}_{DGMCF}^{t}-\boldsymbol{J}_{DGMCF}^{t-1}\right\|<\varepsilon$ 被满足，停止迭代,并记 $\boldsymbol{X}_{l+1}$=$\boldsymbol{V}_l$；

8：end for

输出：分解后的矩阵 $\boldsymbol{W}=\prod_{l=1}^{L}\boldsymbol{W}_l$ 和 $\boldsymbol{V}=\boldsymbol{V}_L$。

2.6 DGMCF 算法复杂度分析

为了比较本文所提算法和其他算法的计算复杂度，本文对 CF、GCF、GMCF 和 DGMCF 算法的计算复杂度进行了比对。经过 t 次迭代更新后，CF 和 GCF 算法的计算复杂度分别为 $O(tN^2K+N^2M)$ 和 $O(N^2M+NM^2+tN^2K)$。对于 GMCF 和 DGMCF 算法，均对数据进行了 L 层的分解，GMCF 算法构建 p 最近邻图的计算复杂度为 $O(N^2M+N^2p)$，DGMCF 算法创建数据图和特征图所需的计算复杂度为 $O(N^2M+NM^2)$，所以 GMCF 和 DGMCF 算法总的计算复杂度分别为 $O(L(N^2M+N^2p+tN^2K))$，$O(L(N^2M+NM^2+tN^2K))$。

3 数值实验

3.1 聚类实验

为了验证本文所提算法的有效性，分别在文本数据库 TDT2 和图像数据库 PIE、COIL20 三个数据集上将所提 DGMCF 算法与 CF、LCCF、MCF、GMCF 和 GCF 算法进行了聚类比较实验。聚类实验中常用准确率(accuracy, ACC)和归一化互信息(normalized mutual information, NMI)[5]作为聚类算法的评价标准。

- TDT2 文本数据库：包含了 56 类的 10 021 个文档。
- PIE 人脸数据库：选择了 68 人的 11 554 幅图像，除第 38 个人 164 幅图像外，其余人都是 170 幅图像。本实验中图像的大小为 32*32 的灰度图像。
- COIL20 物体图像数据库：包含了 20 个物体的 1 440 幅图像，图像的大小为 32*32 的灰度图像。

实验中，对于 MCF、GMCF 及 DGMCF 方法结合文献[],选择 L=10，并且每一层的最大迭代次数设置为 500 次。对于 LCCF、GMCF、GCF 和 DGMCF 方法，构造图所需近邻点数 p 都设置为 5。对不同的聚类数 k $(k=2,\cdots,10)$，DGMCF 算法与五种相关的比较算法在三个数据库上 20 次实验平均的聚类结果如表 1～3 所示。从表 1～3 可得到如下的结论：

a) 在 TDT2 数据库中，本文所提算法比传统 CF 算法平均 ACC 和 NMI 分别提高了 11.23%和 11.67%，比 GCF 算法分别提高了 1.94%和 2.65%。比 GMCF 算法分别提高了 2.28%和 2.33%。在 PIE 数据库中，DGMCF 比 CF 算法平均 ACC 和 NMI 分别提高了 13.70%和 12.18%，比 GCF 算法分别提高了 3%和 2.45%。比 GMCF 算法分别提高了 3.49%和 2.90%。在 COIL20 数据库中，DGMCF 比 CF 算法平均 ACC 和 NMI 分别提高了 10.93%和 13.18%，比 GCF 算法分别提高了 3.36%和 4.02%。比 GMCF 算法分别提高了 4.02%和 7.17%。

b) LCCF 算法利用了数据分布的几何结构信息，所取得的聚类效果比 CF 本身要好。这就表明数据分布的几何结构在进行聚类工作时是有效的，特别对图像数据集，它们空间分布具有潜在的流形结构。在三个数据库中 GCF 算法所得效果优于 LCCF 算法，这是因为 GCF 算法不但考虑了数据流形的几何结构信息，而且也利用了特征流形的结构信息。另外，GMCF 算

法所取得的效果优于 CF 及 LCCF，主要原因是该方法通过逐层分解并在每一层结合数据分布的几何结构信息，使得该方法能够学习出更利于数据表示的特征。

c) 最后本文提出的 DGMCF 算法比 GCF 算法的聚类质量更好，其主要原因是 DGMCF 算法利用了多层分解的思想，在每一层分解中结合数据流形及特征流形的结构信息，能够更加深层次地挖掘出表征数据最本质的特征，尤其是在图像数据集中所取得的聚类效果比文本数据集要更加的明显些。

表1　TDT2 数据库上的聚类实验结果

k	accuracy/%						normalized mutual information/%					
	CF	LCCF	MCF	GMCF	GCF	DGMCF	CF	LCCF	MCF	GMCF	GCF	DGMCF
2	86.13	94.23	95.34	96.45	96.21	97.31	67.28	84.18	83.49	85.38	85.94	87.79
3	79.67	88.53	87.63	89.53	91.04	93.56	68.91	78.23	80.12	83.43	81.33	84.63
4	78.29	86.65	85.46	87.16	89.45	91.28	69.78	77.58	79.94	81.16	80.64	84.46
5	73.63	83.41	83.62	84.87	87.66	89.63	64.65	72.04	74.36	76.32	76.56	79.36
6	74.82	81.26	81.47	83.47	84.61	85.42	67.86	74.86	75.28	77.73	77.41	80.24
7	70.53	79.12	80.11	80.65	80.39	82.25	66.54	73.32	73.36	75.65	75.17	77.55
8	67.04	74.28	74.87	76.75	75.20	78.32	68.31	69.73	70.23	73.43	72.58	75.76
9	67.35	75.18	75.68	75.78	73.28	75.85	69.43	70.42	69.68	71.65	72.20	74.05
10	68.65	69.24	71.56	71.94	71.86	73.54	69.81	69.59	70.45	71.87	71.85	73.76
Avg.	74.01	81.32	81.75	82.96	83.30	85.24	68.06	74.44	75.21	77.40	77.08	79.73

表2　PIE 数据库上的聚类实验结果

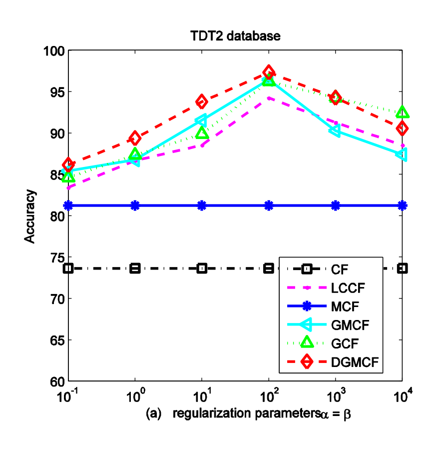
k	accuracy/%						normalized mutual information/%					
	CF	LCCF	MCF	GMCF	GCF	DGMCF	CF	LCCF	MCF	GMCF	GCF	DGMCF
2	57.23	62.89	66.43	68.35	67.14	70.38	48.62	60.43	62.17	64.23	64.59	67.35
3	58.14	58.42	63.54	68.46	69.25	71.43	49.14	57.85	56.12	60.87	61.78	64.87
4	58.36	60.12	64.02	71.24	71.71	72.57	47.23	53.54	55.45	58.63	58.38	61.56
5	58.89	60.61	63.21	71.36	72.86	74.47	48.54	54.42	54.48	57.13	56.36	58.88
6	57.63	59.01	62.53	68.11	69.67	73.54	46.13	51.28	52.87	53.76	54.78	57.34
7	57.80	59.36	61.15	67.32	70.27	72.08	45.64	48.24	49.38	51.35	51.87	54.23
8	55.97	56.12	60.89	65.76	64.03	69.35	41.29	45.11	46.12	48.56	49.41	51.54
9	57.26	57.45	60.43	64.92	65.39	70.18	39.56	42.56	44.78	47.75	48.35	50.05
10	56.85	57.07	61.32	64.53	64.10	67.45	37.45	40.76	42.35	44.85	45.61	47.38
Avg.	57.57	59.01	62.61	67.78	68.27	71.27	44.84	50.47	51.52	54.12	54.57	57.02

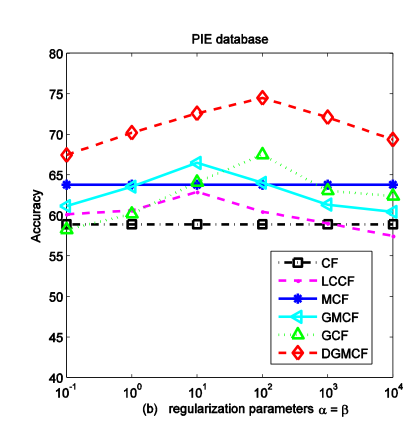
表3　COIL20 数据库上的聚类实验结果

k	accuracy/%						normalized mutual information/%					
	CF	LCCF	MCF	GMCF	GCF	DGMCF	CF	LCCF	MCF	GMCF	GCF	DGMCF
2	89.72	90.74	91.25	91.86	92.48	95.06	71.13	74.51	74.89	76.47	80.40	84.28
3	79.34	84.22	85.03	86.12	85.36	90.31	63.21	68.69	73.02	75.23	76.35	80.45
4	73.04	78.14	79.34	80.28	82.69	87.43	66.38	70.63	72.16	74.26	77.43	81.26
5	71.33	74.46	74.12	78.32	79.23	83.37	67.67	72.22	70.31	72.65	78.56	82.43
6	75.21	79.59	78.75	80.56	82.90	84.52	65.33	68.81	68.63	70.24	74.89	78.37
7	63.85	70.08	74.57	75.27	73.62	76.87	66.67	70.57	69.89	71.32	75.31	79.29
8	64.64	71.64	72.55	73.24	75.51	78.48	67.28	70.67	71.96	72.35	76.45	80.77
9	62.86	67.87	69.13	71.30	70.02	73.29	66.40	69.86	70.21	71.51	72.71	77.32
10	62.15	65.71	66.69	67.37	68.44	71.16	66.27	68.69	69.28	70.38	70.63	74.76
Avg.	71.35	75.83	76.83	78.26	78.92	82.28	66.70	70.52	71.15	72.71	75.86	79.88

3.2 参数选择

DGMCF 算法主要有四个参数：构造对偶图时所需近邻点数 p、两个正则化参数 α 和 β 及分解的层数 L。对于分解的层数 L 的讨论，在文献[17]中已经给出了详细的讨论。为简便起见，设定 $\alpha=\beta$。下面讨论 DGMCF 算法关于参数的稳定性，即给出算法参数设定值与算法的聚类准确率(accuracy)之间的变化情况。当讨论正则化参数 α 时，设定所选择的近邻点数 p =5。类似地，讨论 p 时，设定正则化参数 α =100。为了同时能与 LCCF、GMCF 和 GCF 算法进行比较，实验中也设定 LCCF、GMCF 和 GCF 算法中的图正则参数与近邻点数与本文取相同数值。分别在文本数据库 TDT2 和图像数据库 PIE、COIL20 中进行了比较实验，同时 CF、MCF 算法作为参考数值也一并给出，如图 1、2 所示。





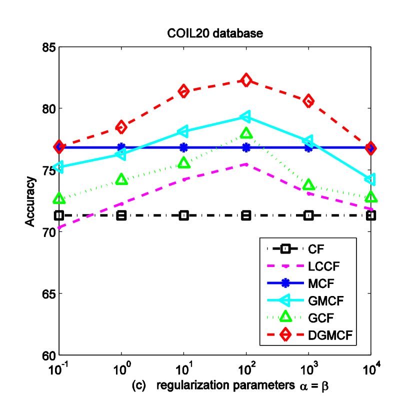
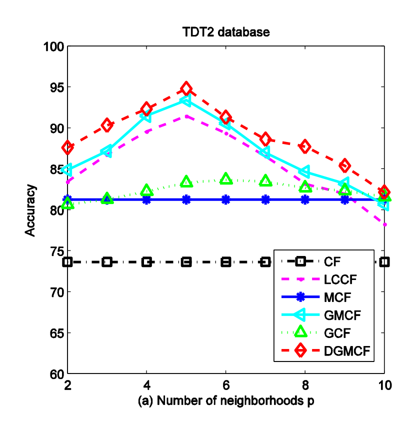
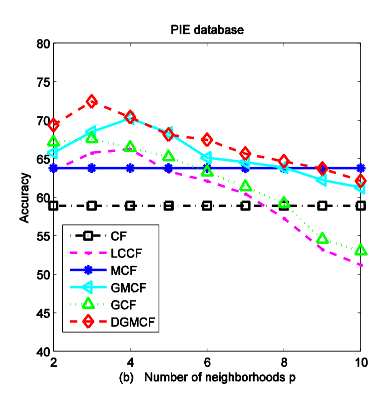


图 1 各种算法的聚类准确率随正则化参数($\alpha=\beta$)改变的变化情况





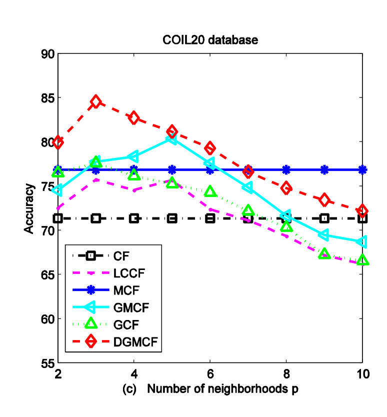


图 2 各种算法的聚类准确率随构建图边数(p)改变的变化情况

由图 1、2 所示的结果，可以得到如下的结论：

a) 提出的 DGMCF 算法对于两个正则参数来讲是非常稳定的。当两个正则化参数取值在 1～1000 时，算法能够取得不错的聚类效果。同时 DHCF 算法本身所得的结果要优于其他算法。

b) 提出的 DGMCF 算法的聚类准确率随着最近邻数取值过大而降低，这是由于过大的最近邻数生成的图不再能准确地反映样本间固有的几何结构。

4 结束语

本文提出了一种基于对偶图正则化的多层概念分解算法，利用多层分解思想，在每一层同时利用数据流形和特征流形的几何结构信息，进一步挖掘出表征数据最本质的特征。本文还给出了 DGMCF 算法的目标函数及迭代更新公式，并给出了算

法的收敛性证明。大量的实验结果表明提出的算法比已有相关算法在数据表示方面的性能更好。

参考文献：

[1] Yi Yugen, Shi Yanjiao, Zhang Huijie, et al. Label propagation based semi-supervised nonnegative matrix factorization for feature extraction [J]. Neurocomputing, 2015, 149: 1021-1037.

[2] 王萌萌，左万利，王英，等. 一种基于加权非负矩阵分解的多维用户人格特质识别算法 [J]. 计算机学报, 2016, 39 (12): 2562-2577.

[3] 贺超波，汤庸，杨阿祧，等. 基于分布式非负矩阵分解的大规模主题社区挖掘 [J]. 中国科学：信息科学, 2016, 46 (6): 714-728.

[4] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization [J]. Nature, 1999, 401: 788-791.

[5] Xu Wei, Gong Yihong. Document clustering by concept factorization [C]// Proc of International Conference on Research and Development in Information Retrieval. 2004: 202-209.

[6] Liu Haifeng, Wu Zhaohui, Li Xuelong et al. Constrained nonnegative matrix factorization for image representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 34 (7): 1299-1311.

[7] Liu Haifeng, Yang Genmao, Wu Zhaohui, et al. Constrained concept factorization for image representation [J]. IEEE Trans on Cybernetics, 2014, 44 (7): 1214-1224.

[8] Tenenbaum J B, Silva V D, Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290 (5500): 2319–2323.

[9] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290 (5500): 2323-2326.

[10] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering [C]// Advances in Neural Information Processing Systems. 2002: 585-591.

[11] Guan Naiyang, Tao Dacheng, Luo Zhigang, et al. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent [J]. IEEE Trans on Image Processing, 2011, 20 (7): 2030-2048.

[12] Cai Deng, He Xiaofei, Han Jiawei, et al. Graph regularized nonnegative matrix factorization for data representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33 (8): 1548-1560.

[13] Cai Deng, He Xiaofei, Han Jiawei. Locally consistent concept factorization for document clustering [J]. IEEE Trans on Knowledge and Data Engineering, 2011, 23 (6): 902-913.

[14] Cichocki A, Zdunek R. Multilayer NMF using projected gradient approaches [J]. International Journal of Neural Systems, 2007, 17 (6): 431-446.

[15] Rajabi R, Ghassemian H. Spectral unmixing of hyperspectral imagery using multilayer NMF [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12 (1): 38-42.

[16] Li Xue, Zhao C, Shu Z, et al. Multilayer concept factorization for data representation [C]// Proc of the 10th International Computer Science & Education. 2015: 486-491.

[17] Li Xue, Shen Xiaobo, Shu Zhenqiu, et al. Graph regularized multilayer concept factorization for data representation [J]. Neurocomputing, 2017, 238: 139-151.

[18] Shang Fanhua, Jiao Lichen, Wang Fei. Graph dual regularization non-negative matrix factorization for co-clustering [J]. Pattern Recognition, 2012, 45 (6): 2237-2250.

[19] Ye Jun, Jin Zhong. Dual-graph regularized concept factorization for clustering [J]. Neurocomputing, 2014, 138 (3) , 120-130.